Researchers are helping artificial intelligence understand fairness

22 March 2021, by Matt Davenport

"We're trying to design AI systems that aren't just for computer science, but also bring value and benefits to society," Tan says. Credit: Creative commons via Pexels

"What is fair?" feels like a rhetorical question. But for Michigan State University's Pang-Ning Tan, it's a question that demands an answer as artificial intelligence systems play a growing role in deciding who gets proper health care, a bank loan or a job.

With funding from Amazon and the National Science Foundation, Tan has been working for the last year to teach artificial intelligence algorithms how to be more fair and recognize when they're being unfair.

"We're trying to design AI systems that aren't just for computer science, but also bring value and benefits to society. So I started thinking about what are the areas that are really challenging to society right now," said Tan, a professor in MSU's Department of Computer Science and Engineering.

"Fairness is a very big issue, especially as we become more reliant on AI for everyday needs, like health care, but also things that seem mundane, like spam filtering or putting stories in your news feed."

As Tan mentioned, people already trust AI in a variety of applications and the consequences of unfair algorithms can be profound.

For example, investigations have revealed that AI systems have made it harder for Black patients to access health care resources. And Amazon scrapped an AI recruiting tool that penalized job applicants who were women in favor of men.

Tan's research team is contending with such problems on multiple fronts. The Spartans are looking at how people use data to teach their algorithms. They're also investigating ways to give algorithms access to more diverse information when making decisions and recommendations. And their work with the NSF and Amazon is attempting to broaden the way fairness has usually been defined for AI systems.

A conventional definition would look at fairness from the perspective of an individual; that is, whether one person would see a particular outcome as fair or unfair. It's a sensible start, but it also opens the door for conflicting or even contradictory definitions, Tan said. What's fair to one person can be unfair to another.

So Tan and his research team are borrowing ideas from social science to build a definition that includes perspectives from groups of people.

"We're trying to make AI aware of fairness and to do that, you need to tell it what is fair. But how do you design a measure of fairness that is acceptable to all," Tan said. "We're looking at how does a decision affect not only individuals, but their communities and social circles as well."

Consider this simple example: Three friends with identical credit scores apply for loans worth the same amount of money from the same bank. If the

bank approves or denies everyone, the friends would perceive that as more fair than a case where only one person is approved or denied. That could indicate that the bank used extraneous factors that the friends might deem unjust.

Tan's team is building a way to essentially score or quantify the fairness of different outcomes so AI algorithms can identify the most fair options.

Of course, the [real world](#) is much more complex than this example, and Tan is the first to admit that defining fairness for AI is easier said than done. But he has help—including from the chair of his department at MSU, Abdol-Hossein Esfahanian.

Esfahanian is an expert in a field known as applied graph theory that helps model connections and relationships. He also loves learning about related fields in computer science and has been known to sit in on classes taught by his colleagues, as long as they're comfortable having him there.

"Our faculty are fantastic in imparting knowledge," Esfahanian said. "I needed to learn more about data mining, and so I sat in one Dr. Tan's courses for a semester. From that point on, we started communicating about research problems."

Now, Esfahanian is a co-investigator on the NSF and Amazon grant.

"Algorithms are created by people and people typically have biases, so those biases seep in," he said. "We want to have fairness everywhere, and we want to have a better understanding of how to evaluate it."

The team is making progress on that front. This past November, they presented their work at an online meeting organized by NSF and Amazon as well as at a virtual international conference hosted by the Institute of Electrical and Electronics Engineers.

Both Tan and Esfahanian said the community—and the funders—are excited by the Spartans' progress. But both researchers also acknowledged that they're just getting started.

"This is very much ongoing research. There are a lot of issues and challenges. How do you define fairness? How can you help people trust these systems that we use every day?" Tan said. "Our job as researchers is to come up with solutions to these problems."

Provided by Michigan State University